班　　级：信研 2305
学　　号：2023200820



北京化工大学

深度学习课程作业

题　目　不良光照条件对深度学习目标检测的影响研究

专　　业　计算机科学与技术

学　　生　刘景硕

课程教师　李瑞瑞

2024 年 1 月 6 日

摘 要

不良光照条件下的目标检测是一个重要的图像处理任务，目前的研究主要通过图像增强来减少图像噪声，同时改进网络结构和数据集来适应不良光照条件下的目标检测。然而，很少有人研究不良光照条件对目标检测的具体影响。因此在本文中，我们通过算法生成模拟不良光照条件的数据集，在不同的噪声条件下进行目标检测，统计检测结果，对影响进行研究。

关键词：低照度，YOLOv5，目标检测

一、研究背景

1.1 研究意义

近年来，自动驾驶逐渐兴起并受到诸多研究者的青睐。在自动驾驶领域中，目标检测对安全驾驶起到重要作用，2D 目标检测通过接收相机的数据，预测出周围存在的物体和交通标志、斑马线等图形信息。随着深度学习的发展，2D 目标检测的精度和速度在不断地提高，目前的目标检测算法主要有 One-Stage 和 Two-Stage 两类，代表性的算法有 YOLO 系列、RCNN 系列、Retina-Net 等。正常光照条件下，2D 目标检测已经能够达到很高的精度。

然而在现实中，环境的复杂性使得图像的输入质量并不能得到保证，不良光照条件下，输入图像中的特征信息会变得模糊，这一变化会对目标检测的性能造成影响。而在自动驾驶领域，为了实现智能系统的全天候运行，不良光照条件下的目标检测是一个必须要解决的问题。

有关不良光照条件下的目标检测工作已经有很多，许多研究尝试通过图像增强和前处理方法减少不良光照条件下图像中的噪声，包括调整图像对比度、增强亮度、去除噪声等手段，使目标更容易被检测到。另外，近些年深度学习技术在解决不良光照条件下的目标检测问题上取得了显著的进展，深度学习模型能够从大量数据中学习特征，对光照变化具有一定的鲁棒性。同时，一些研究将可见光图像与其他传感器，如红外、热成像的数据进行融合，以提高在不同光照条件下目标检测的性能，多模态融合有助于提供更全面可靠的信息。

然而，很少有研究者分析不良光照条件对目标检测的影响大小。因此需要量化不良光照条件对目标检测的影响，这一研究可以让科研工作者更加直观的了解图像噪声对目标检测的影响，从而帮助研究人员针对不同场景的需要更加合理的制定目标检测的改进策略。

1.2 研究现状

对于目标检测的方法，目前有 One-Stage 和 Two-Stage 两个类别，有许多优秀的目标检测模型。One-Stage 的方法直接提取特征，然后进行目标检测，速度快，能够避免背景的误报，但目标检出率和定位精度较低，主要算法有：YOLO 系列，SSD 系列，Retina-Net，DetectNet 等；Two-Stage 的方法则先生成候选

区域，然后进行目标分类与定位，训练时间长，速度慢，但目标检测精度高，主要算法有：RCNN 系列，SPPNet，R-FCN 等。

针对不良光照目标检测，目前解决该问题主要是利用低照度图像增强算法对图像进行处理，常用的图像增强方法有直方图均衡化，同态滤波，Retinex 理论等，同时随着深度学习的发展，基于深度学习的图像增强方法也逐渐得到了完善。通过改进目标检测网络和使用不良光照数据集的方式，也能够提升不良光照条件下的目标检测能力。

二、相关理论和技术

2.1 目标检测算法对比

2.1.1 RCNN 算法

RCNN 是一种基于卷积神经网络的两阶段目标检测算法，它是第一个成功应用于目标检测的卷积神经网络，RCNN 的核心思想是对候选区域进行分类，首先使用筛选算法提取出候选区域，通过卷积神经网络对每个候选区域进行特征提取，之后使用支持向量机对每个候选区域进行分类，对于分类为正样本的候选框，使用回归器进行微调，以得到更准确的检测框。相比于传统算法，RCNN 具有更高的准确率，但是计算速度较慢，并且会占用大量的计算资源，而且传统的卷积神经网络需要固定尺寸的输入，这会导致图像在缩放过程中丢失一部分信息，为了解决这个问题，人们提出了 Spatial Pyramid Pooling Networks (SPPnet)结构，即空间金字塔池化，使得图像不再需要固定的输入尺寸。

为了提高 RCNN 的检测速度，作者在 RCNN 的基础上提出了 Fast-RCNN，相比于 RCNN 中每个候选区单独进行特征提取，Fast-RCNN 以整张图片作为输入进行特征提取，大大减小了计算量。为了进一步减小计算量，人们在此基础上引入了 Region Proposal Network，RPN 为全卷积网络，将候选框的提取和目标预测集成到一起，进一步减小了计算量，同时提高了网络性能。

2.1.2 SSD 算法

SSD 算法是一种单阶段的目标检测算法，因此是一种快速的目标检测算法。SSD 的网络由 VGG16 网络和特征提取层构成。SSD 主要使用了两个技术，一是多尺度预测，在多层神经网络中，从不同尺度的特征图中检测目标，从而处理图

像中不同大小的物体，提高检出率，二是SSD算法中会预测出多个具有不同长宽的预测框，并存储这些预测框，通过置信度得分和非极大值抑制对预测框进行调整，筛选出最终的预测框并确定其类别。SSD网络比较简单，该算法将目标检测整合到一个网络中，加快了检测速度，而多尺度预测则提高了算法的性能，同时训练时只需要标注检测框和类别，不需要繁琐的先验框设置，这使得SSD算法易于人们的训练，并能够集成到其他的网络中，因此SSD算法得到了广泛的应用。

2.1.3 YOLO算法

Redmon等人提出了YOLO算法，全名为You Only Look Once，这是一种速度很快的单阶段目标检测算法，它可以在保证检测性能的同时实现很快的检测速度。YOLO算法将目标检测问题转化为回归问题，通过单个神经网络对图像直接进行目标检测。YOLO算法首先将图像划分为S×S的网格，对每个网格进行计算，预测该网格中是否包含物体，如果一个物体的中心落到了这个网格中，就由这个网格负责预测该物体的类别、位置、大小，每个网格预测检测框和检测框的得分，并通过非极大值抑制进行处理，排除重复预测的物体。YOLO算法结构简单，运行速度快，能够满足实时性检测的要求，但也因此会出现一些偏差，同时因为其划分网格的设计，由于网格数量和每个网格能够产生的预测框的数量有限，对于图像中出现的密集小物体的预测性能较差。

综合考虑目标检测的性能和速度，后续实验采用YoloV5目标检测算法，其网络结构如图1所示。

2.2YoloV5网络概述

在输入端上，YOLOV5采用了Mosaic数据增强，采用四张图片，通过随机缩放，随机裁剪，随机排布的方式进行拼接，丰富了数据集，同时对于小目标的随机缩放使得算法对小目标的检测效果得到了提升，提高了网络模型的鲁棒性，同时这种将四张图片进行拼接的方式，使得模型在训练时能够同时输入四张图片的数据，在训练时一个GPU也能够达到较好的训练效果

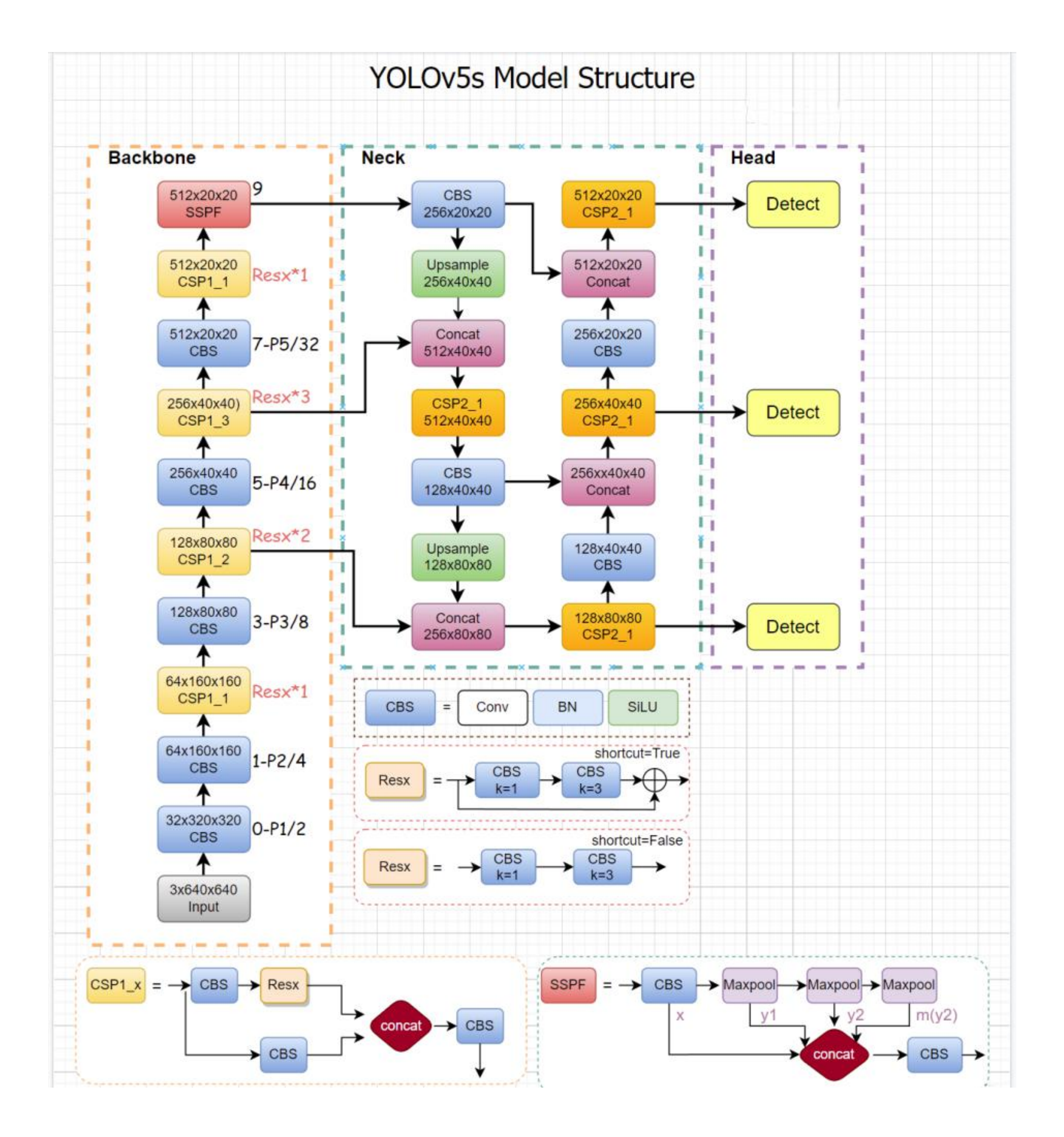


图 1 网络结构

YOLOV5 的 Backbone 中采用了 Focus 模块，在 Focus 中采用了切片结构，减小了图片的尺寸，增加了通道数，切片操作如图 2 所示，采用该操作，图片尺寸减少了一半，通道数则变成了四倍，达到了类似下采样的目的，同时图片中的信息没有丢失。另一个模块是 CSP 结构，CSP 结构用于解决计算量问题，优化了网络中的计算，在保证精度的同时减少了计算量。

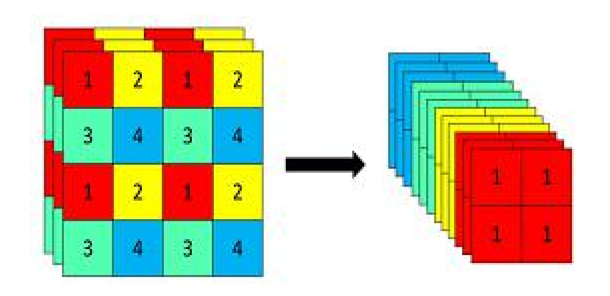


图 2 切片操作

整体结构类似 FPN+PAN 结构，在目标检测中，深层的特征图带有较强的语义信息，但是位置信息较弱，而浅层的特征图中，位置信息较强，但是携带的语义信息较弱，FPN 结构则将深层的语义信息传递到浅层，PAN 结构则将浅层的位置信息传递到深层，从而增强网络在多个尺度上的语义表达和物体定位能力，进一步提高了特征提取能力。

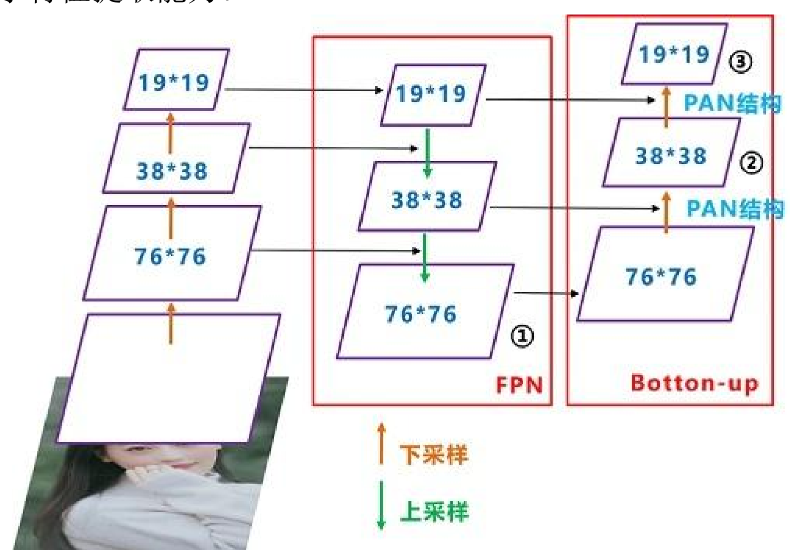


图 3 FPN+PAN

2.3 算法输入和输出

目标检测算法的输入一般为 RGB 图像，算法需要预测出物体的类别和在图像中的位置，因此输出一般由物体类别和检测框的表示信息构成。目标类别由置信度来确定，算法会选择所有预测类别中置信度最大的一类作为最终的类别，同时在网络中会设置置信度阈值，超过阈值的物体才会作为正样本，低于阈值的作为负样本。检测框的位置描述一般由某一顶点或者中心点的坐标加上检测框的长宽来描述其位置和大小。

三、不良光照图像的生成

来自摄像头的图像输入易受到光照和天气的影响，在不良光照的影响下，图像会产生不同程度的失真。现实环境中获取不良光照图像难度较大，且无法获得原图-不良光照图像的图像对，但不良光照图像可以通过图像处理算法来生成，且噪声的强度可以通过参数来量化。因此可以通过现有的图像处理算法，获取不良光照条件下的图像。

强光、弱光图像生成：

计算机中常用的是 RGB 色彩模式，通过将 RGB 色彩模式转变为 HSV 色彩模式，即色相(Hue)、饱和度(Saturation)和亮度(Value)，改变其中的亮度分量来控制图像中的光照强度，再将图像转换回 RGB 色彩模式，就能够得到我们想要的任意光照强度的数据集。转换公式如下：

$$V = MAX(R, G, B)$$

$$S = \begin{cases} \dfrac{V - MIN(R, G, B)}{V} & V \neq 0 \\ 0 & else \end{cases}$$

$$H = \begin{cases} \dfrac{60(G - B)}{V - MIN(R, G, B)} & V = R \\ 120 + \dfrac{60(B - R)}{V - MIN(R, G, B)} & V = G \\ 240 + \dfrac{60(R - G)}{V - MIN(R, G, B)} & V = B \end{cases}$$

$$if \quad H < 0, H = H + 360$$

基于光学模型的雾霾图像生成：

雾霾图像和原始图像的数学关系可以表示为：

$$I(x) = J(x)t(x) + L(1 - t(x))$$

公式中，I(x)为有雾图像，x 为像素的坐标值，J(x)为原始图像，L 为全球大气光成分，t(x)为透射率，在雾霾场景中，由于雾的影响物体的反射能量 J(x)t(x)被削弱，同时雾霾引起的大气散射 L(1-t(x))增加，降低图像饱和度。由此，有雾图像可以看作由物体的反射光和大气的散射光形成。

设在(x, y)生成雾霾，中心为(X, Y)，size 为雾化尺寸，通过引入参数α，控制雾霾浓度，雾霾的合成公式如下：

$$d = -0.04 \times \sqrt{(x - X)^2 + (y - Y)^2} + size$$

$$td = e^{-\alpha \times d}$$

$$I(x, y) = J(x, y) \times td + A \times (1 - td)$$

这种策略可以通过控制算法参数来量化图像中的噪声大小，同时生成的原图和一系列相应的不良光照图像能够用于对比试验，便于开展后续的研究。

四、实验和分析

4.1 模型训练

使用 VOC2007 数据集对模型进行训练，总共训练 50 个 epoch，训练中设置 batch-size 为 16，首先以 0.01 的初始学习率训练 10 个 epoch，随后，调整学习率为 0.001，继续训练 20 个 epoch，最终，调整学习率为 0.0001，继续训练 20 个 epoch，至此，训练中的各参数变化已不明显，模型基本收敛。

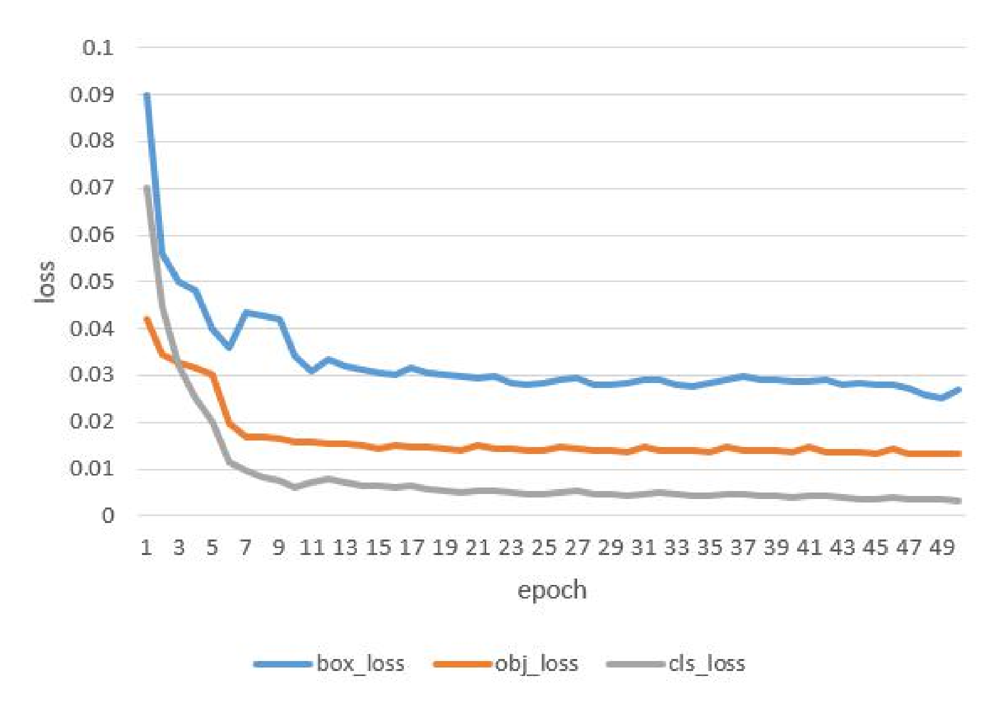


图 4 训练中模型 loss 变化

模型在测试集中的表现如表 1 所示：

表 1 模型性能

Class	Precision	Recall	mAP	mAP0.5:0.95
Bicycle	0.866	0.847	0.905	0.587
Bus	0.881	0.781	0.861	0.646
Car	0.801	0.837	0.887	0.635
Motorbike	0.832	0.836	0.879	0.567
Person	0.834	0.862	0.898	0.572

4.2 噪声影响分析

使用训练好的 YOLOV5 目标检测算法对不良光照图像进行目标检测，在检测中设置置信度阈值为 0.3，目标检测置信度大于 0.3 的会被识别为正样本。通过控制图像处理算法的参数，在强光、弱光、雾霾条件下分别对图像进行检测，统计同一物体在不同噪声等级中对应的目标检测置信度，通过对检测结果进行整理，得到了不同噪声下图像中物体的目标检测置信度变化曲线。

强光条件下的统计结果如图 5 所示，总体上来看，噪声等级较低时，大部分物体的置信度变化较小。在噪声等级在 1 到 10 级之间，物体 3，物体 5，物体 8 的置信度出现了明显下降，低于设置的阈值，其他物体的目标检测置信度变化较小，当噪声等级继续变高时，物体 1 和物体 2 的置信度也出现下降。

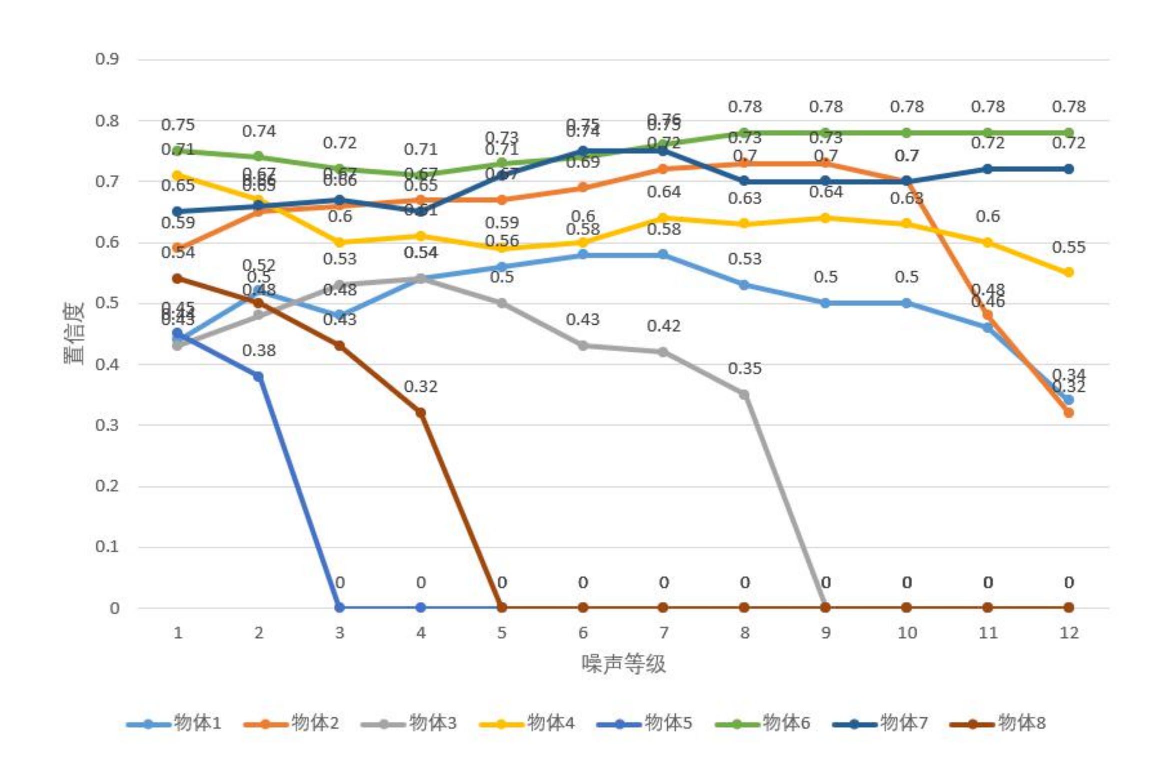


图 5 强光条件下噪声-置信度变化

如图 6 所示，与强光条件相比，弱光条件下的目标检测置信度相对平稳，大部分物体的置信度几乎不受影响，同时由于图像中的反光较弱，部分物体出现置信度增加的情况。

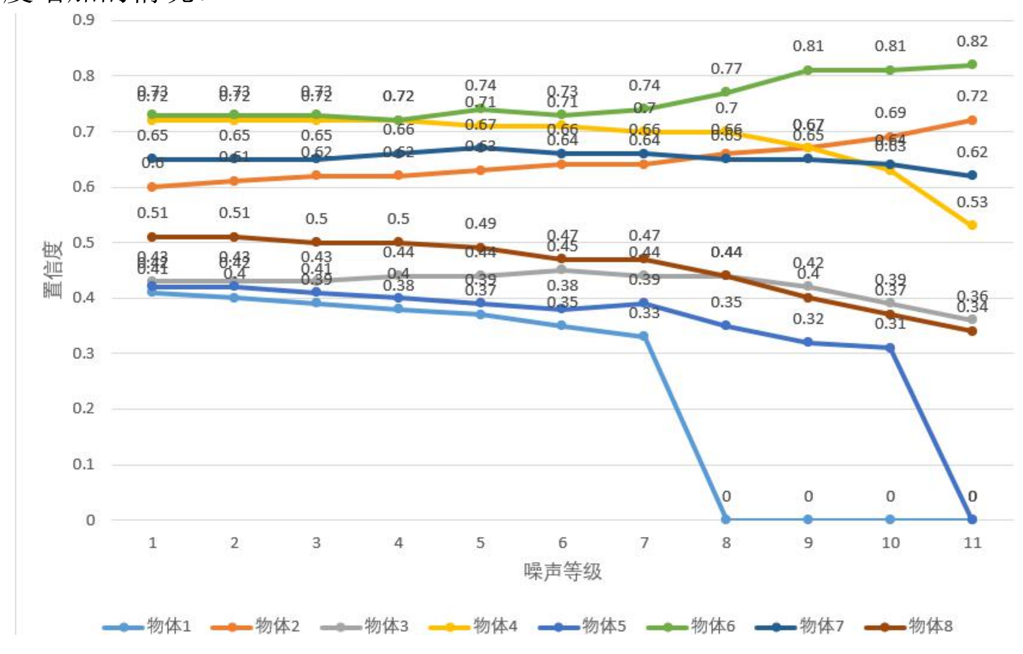


图 6 弱光条件下噪声-置信度变化

雾霾条件下，大部分物体都受到较大的影响，少部分物体由于生成的雾霾浓度较低，受影响程度较轻，如图 7 所示。

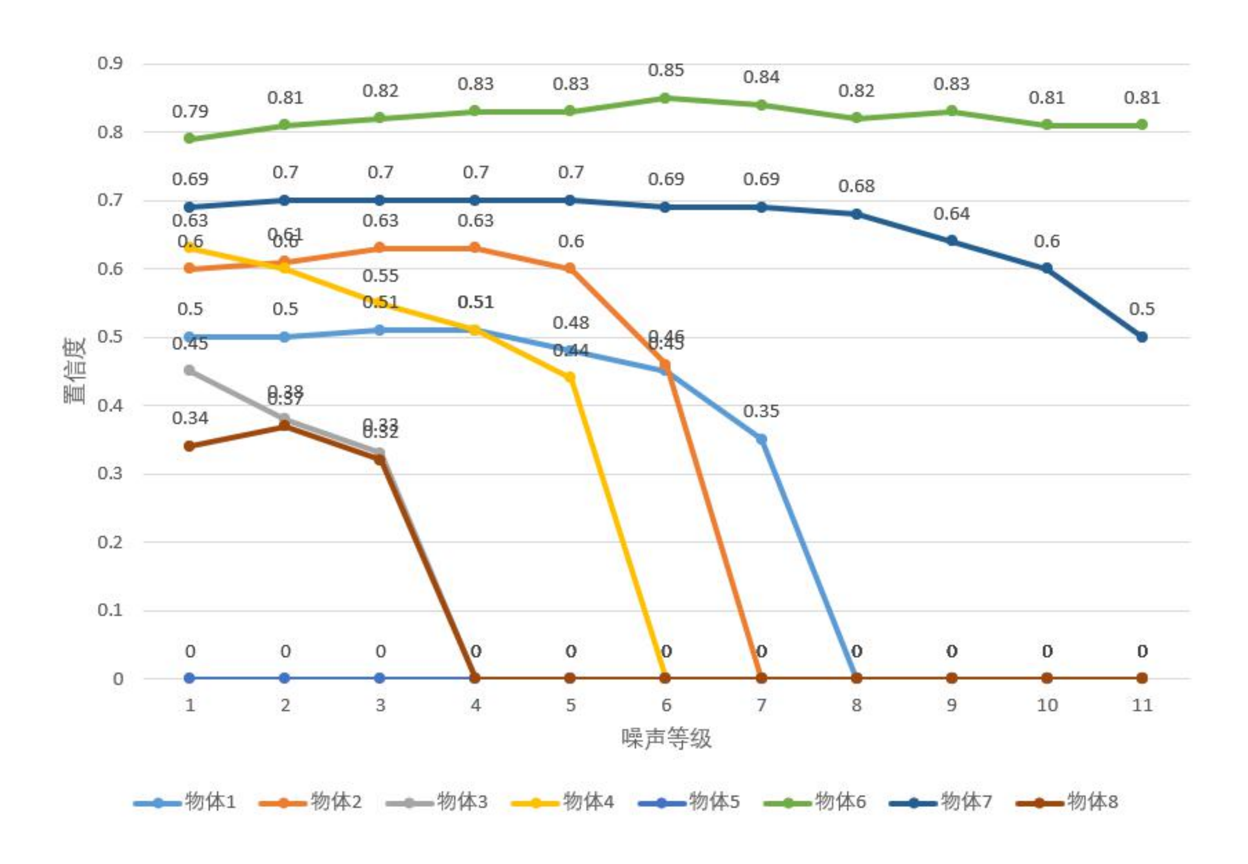
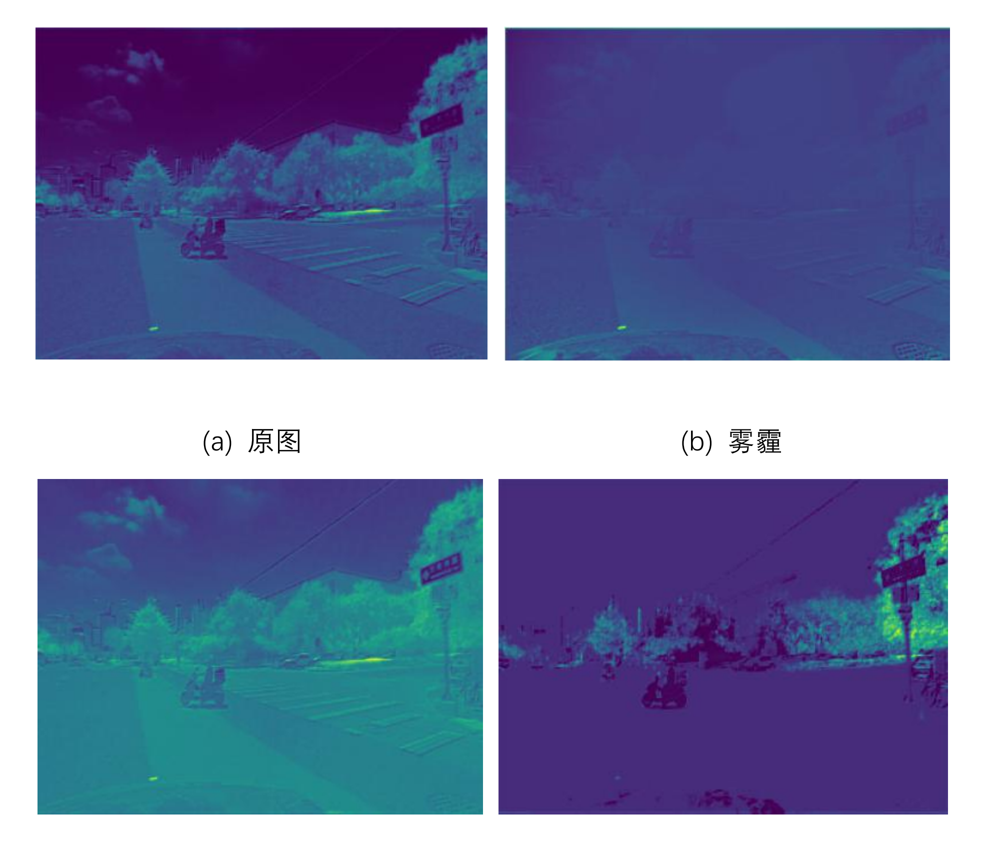


图 7 雾霾条件下噪声-置信度变化

通过特征图可视化，输出原图的浅层特征图和不同噪声条件下的浅层特征图，结果如图 8 所示：



(a) 原图

(b) 雾霾

(c) 暗光　　　　　　　　　　　　(d) 强光

图 8 浅层特征图

噪声的引入使图像对比度下降，物体的特征变的模糊，增加了算法识别物体特征的难度，影响目标检测的结果。三种噪声条件下，弱光对特征的影响较小，强光和雾霾对特征的影响较大，同时由于物体自身因素不同，受到影响的程度并不相同。小物体和存在遮挡的物体由于包含的可识别信息较少，因此在噪声的影响下置信度会很快的低于阈值，同时，与背景对比不明显的物体也容易在噪声影响下产生误判。

五、总结

通过量化噪声的大小、控制噪声等级的方式，本实验在强光、弱光和雾霾三种噪声场景下，统计不同噪声等级的目标检测置信度，并分析统计的结果。随着图像中噪声的增加，特征图中物体被误判的几率增加，影响目标检测的结果。强光条件下，在噪声等级达到 10 之后，大部分物体的置信度都会受到影响；弱光条件下，置信度的变化较小，在噪声等级达到 11 之后，少部分物体的置信度会出现变化；雾霾条件下，在噪声等级达到 8 之后，大部分物体置信度都会出现明显下降。同时噪声对不同物体的影响并不均衡，由于自身因素不同，受到影响的程度并不相同。

然而也要注意到分析所使用的数据是模拟生成的，虽然最后使用现实场景的数据进行了验证，但模拟数据和现实数据依然有出入，由于计算资源不足，所训练的网络检测能力也弱于在服务器上训练的网络。